DOCUMENT RESUME

ED 363 129 FL 021 602

AUTHOR	Gilfert, Susan; Harada, Kunihiko
TITLE	Two Composition Scoring Methods: The Analytic vs. Holistic Method.
PUB DATE	92
NOTE	8p.; The Bulletin is produced by Hokuriku University, Japan.
PUB TYPE	Reports - Research/Technical (143) -- Journal Articles (080)
JOURNAL CIT	Bulletin of Faculty of Foreign Languages; v1 p17-22 1992
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Students; Comparative Analysis; *English (Second Language); Foreign Countries; Higher Education; *Holistic Evaluation; *Scoring; *Testing; Test Reliability; *Writing (Composition)
IDENTIFIERS	*Analytic Approach; Japan

ABSTRACT

This paper addresses the evaluation of the testing of writing of foreign languages and compares two evaluation methods: the analytical method and the holistic method. The analytical method focuses on the mechanics of writing where the writer is measured against a set of empirical standards and a composition is dissected for the critical points. The holistic method looks at a composition as a writing sample and compares the communicative effectiveness of one composition against another of the same type. The writing compositions of 10 college-age English-as-Second-Language (ESL) students were graded by ESL teachers in Japan, some using the holistic method and some using the analytic method. Results showed that both ratings were very close, with the maximum difference of three points in a 20-point scale, and the ratings had a high correlation with the writers' Test of English as a Foreign Language scores. Findings indicate that written English can be tested and the testing of written English can be carried out with satisfactorily high reliability. In terms of measurement, reliability includes the correlation of writing scores with external and internal measures. (Contains 13 references.) (JP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 363 129

Two Composition Scoring Methods: The Analytic vs. Holistic Method

Susan GILFERT
Hokuriku University Kanazawa
Kunihiko HARADA
Georgetown University, School of Languages and Linguistics

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Susan H. Gilfert

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

FL021602

北陸大学外国語学部紀要
第1号 抜刷 1992

Bulletin of Faculty of Foreign Languages
Hokuriku University
Vol. 1 (1992) pp. 17~22

Two Composition Scoring Methods: The Analytic vs. Holistic Method

Susan GILFERT
Hokuriku University Kanazawa
Kunihiko HARADA
Georgetown University, School of Languages and Linguistics

Received October 23, 1992

1. Introduction

In the teaching of foreign languages, many teachers consider that there is a natural progression of skills to be taught: listening, speaking reading, and finally writing. It should be noted that oral skills [listening/speaking] are taught before graphic skills [reading/writing], and passive skills [listening/reading] before active skills [speaking/writing]. After teaching a set of points of a particular skill, it is only natural that the student should be evaluated by means of a test, to measure the proficiency that the student can demonstrate. But what can be tested? How can it be measured? Once the testable material is determined, then the test proceeds. Should the student not demonstrate sufficient proficiency, this is a sign to both the student and teacher that revision is in order. Testing of these various language skills has been under evaluation as long as the tests themselves have been used. (See Allen & Davies, 1977; Carroll, 1980; Kono, 1978; Madsen, 1983 among others) Why? Students are concerned that they are being evaluated by fair and equal tests. Frequently, students gain something valuable to them aterial (a promotion, admission to an academic institution, advancement in their class...) by "passing the test" (i.e., demonstrating sufficient proficiency of the tested material). Students wish to compare their skills and abilities under similar circumstances.

This paper will focus on the evaluation of the testing of the last-taught skill, that of writing. The last skill to be taught is frequently that which is the least understood; most foreign-language learners have problems with expressing themselves adequately in writing. The testing of writing skills has equally been less understood than tests of other skills; what can be tested? How can it be measured? How can it be fair?

---

* 英米語学科
Department of English

2. What Does a Test Test? How Does a Test Test?

In the evaluation of the student, tests are constructed to reflect what and how the student has been taught. The most efficient form of testing is to provide the student with a prompt and request a response; in other words, ask a question and see how the student answers. A test of speaking ability is frequently done by having the student speak; a test of listening is best demonstrated by having the student listen to a set of directions and carry them out, and so on. In the testing of writing, the student is given some prompt which requires her to carefully explain her thinking about some controversial point. Writing within a certain time limit forces the student to quickly but carefully compose her thoughts into the most cohesive form she can demonstrate. How can this test be fair to all testees? There are two widely variant but equally widely used methods of evaluating writing: the analytical method and the holistic method. Each has its strong and weak points as described below.

3. The Analytical Method

Objectiveness and standardization in testing anything are critical, so the examiners discuss and decide on the specific criteria by which each composition is graded before the test is taken by the students. In grading the composition, the examiner assigns points to each criterion and adds all the points to get the overall score. Multiple examiners, reading each paper more than once, are strongly recommended to achieve high reliability. In this study, the following criteria are used:

Criteria	Points
Sentence structure	1 2 3 4 5
the student should show a variety and maturity in the writing	
Grammar	1 2 3 4 5
the student should use acceptable forms and word order	
Vocabulary	1 2 3 4 5
the student should use a variety and preciseness of vocabulary	
Content	1 2 3 4 5
the student should show understanding of her subject	

total score _____

The analytical method tends to focus on the mechanics of writing, not how the writer expresses her thoughts. The student/writer is measured against a set of empirical standards. In the analytical method, a composition is dissected for the critical points. Furthermore, the analytical method pre-determines criteria for grading. Many local tests of writing are graded in this manner.

4. The Holistic Method

This method is used after all the tests have been written. First, multiple examiners skim a sample of several compositions just to get an idea of the level of the students. Then the examiners start grading each composition based on its overall effectiveness as a means to communicate in relation to other compositions. They must use a common scale[1] (as suggested below) and make sure that each scale point, and preferably some intermediate points, are represented in the rating.

4 = one of the best compositions.
3 = good, but not one of the best.
2 = somewhat below the group average.
1 = one of the weakest compositions.

The holistic method looks at the composition as a writing sample, and compares the communicative effectiveness of one composition against another of the same type. The student/writer is measured against her peers. The criteria for grading are not predetermined, but are detemined by the population taking the test. The standardized Test of Written English from ETS in Princeton, NJ and many ESL university-prep programs follow this method of grading writing tests.

5. Experiment

Ten college-age EFL students in the U.S. were asked to write about a festival or a holiday in their countries to inform a friend in the U.S. They were given 30 minutes for this task. Twelve ESL teachers in a M.A. training program in Japan graded these compositions with the analytical method; ten other teachers-in-training in the same program graded these compositions with the holistic method. The analytical ratings were averaged, while the holistic ratings were pooled. The results showed that both ratings were very close, with the maximum difference of 3 points in a 20-point scale; furthermore, the ratings had a high correlation with the writers' TOEFL scorres.

6. Discussion

The analytical method may be easier for many teachers or scorers to grade because they know what to look for in the compositions. All criteria are already agreed upon, and they are all weighed equally. Once points for all criteria are given, the scorer can simply add them up to get the total score. In this study, however, two criteria in the analytical method [Sentence Structure and Grammar] may have been confusing to some scorers. The directions say, for Sentence Structure, "the student should show a variety and maturity in the writing." Also, for Grammar, "the students should use acceptable forms

[1]This scale is more-or-less equivalent to the United States' common A, B, C, D grading scale.

and word order." However, if one looks at the following sentences from one student's composition, it is difficult to decide to which criterion they belong:

- After the soldiers follow the tanks, and artillery.
- When the entire Army has gone by, is the chance for the school.

Do these belong to Sentence Structure or Grammar? The authors suggest using the terms Mechanics rather than Sentence Structure.

On the other hand, the holistic method tells the scorers to judge the composition on the basis of its overall impression using as the single criterion the writer's communicative competence in this one sample of writing. However, what is communicative competence? Its definition surely changes depending on the individuals evaluating the composition. Even within individuals, it is difficult to keep the standard when grading many papers. Actually, it would be difficult to grade compositions without referring back to the already graded papers to make sure which one is betterat communicationg.

The Test of Written English and some ESL preuniversity groups use the holistic grading procedure for determining proficiency levels of ESL students. Composition tests are given to groups of 200 or more students. After the test, but before the grading begins, the readers (usually 5 or more teachers) are all agreed on what is an "4" paper, a "3" paper, a "2" paper, and so on, from random samples and their own experience as graders of ESL written work. Then the papers are divided into batches of about 10 or 15 papers. Each reader reads a paper from the batch and assigns a grade to the paper. When the reading of that batch is finished, the reader takes a short break, then returns to read another batch of papers. Each paper gets read at least twice, by two different readers. The coordinator for the group of tests then goes through the papers and determines the correlation of grades for each paper. Any paper that receives more than 0.5 marks difference from the two readers is read again by a third reader, who determines the mark for the paper. This procedure sounds rather lengthy, but 250 papers can be graded in about 5 hours, providing all the readers persevere. More readers lessens the amount of time.

7. Applications to Japan

Traditionally, the holistic approach has been used by most Japanese teachers if any amount of writing in English were to be done. The holistic method basically uses the same principles as the Monbusho (Ministry of Education) recommends. Students' abilities are curved within the class population and the number of grades to be given are predetermined by percentage. All marks and grades are internally accredited. This is a problem. If a student happens to be in a uniformly good class, one point might make a difference; it is also very difficult for a teacher to award grades.

The problem is exacerbated in the case of a uniformly low class. Some students may get high grades by a very narrow margin, but they are not aware of their abilities in

relation to an exterior criterion such as the Test of Written English. It is a disservice to students to give them a false impression that they are good when they are not.

Japanese students are not taught explicitly to develop composition structure in English. Many ESL students follow the writing structure of their own native literature; English is by no means a universal. (Kaplan, 1966) When Japanese students attempt an external measure such as the Test of Written English, they frequently score poorly until they learn the English pattern of development.

8. Conclusion

The question remains: What can be tested? How can it be measured? How can a test, especially a standardized test such as the Test of Written English, be fair? There are both positive and negative points possible from this study. A positive point is that written English can be tested. The testing of written English can be carried out with satisfactorily high reliability. As to the question of measurement, "reliability" includes the correlation of writing scores with external measures such as the TOEFL, and internal measures such as the teacher's evaluation based on classroom observation and daily interaction with the EFL student. This is a positive point for EFL students. A negative point is that not all students are taught to write in the same rhetorical modes and ways of development. An ESL student from a non-Western background, for example, would have different modes of development from a Western student. Those modes and rhetorical development must be taught in the ESL classroom.

From teaching in Japanese classrooms for many years, the authors have found that Japanese students take many things for granted due to the almost homogeneous society and the socially accepted norms that everyone should do the same thing (i.e., learn the same material at the same rate) at the same time. It is not necessary to speak your opinion aloud, or write it clearly in this case, because the audience is supposed to "get" your subtle message. The subtlety is a virtue in Japan. This affects the organization of a paragraph and an entire essay. Thus, teaching rhetorical modes is strongly recommended, far successful communication in English.

The Authors

Susan Gilfert, M.A. is a Professor at Hokuriku University, Kanazawa. She has been teaching ESL for more than ten years in Japan, Malaysia and the U.S. She is co-author of The Heinemann TOEIC Preparation Book and is interested in standardized testing.
Kunihiko Harada, M.A., is a Senior Lecturer at Trident School of Languages, Nagoya. He has been teaching ESL for more than seven years in Japan and has taught American Literature in the U.S. He is currently doing research in CALL and testing.

Bibliography

Allen, J.P.B. & Davies, A. (Eds.). (1977). *Testing and experimental methods.* London: Oxford University Press.

Carroll, B.J. (1980). *Testing communicative performance: An interim study.* Oxford: Pergamon Press.

Gwin, T. (1991, July). Giving students the write idea: A way to provide feedback on writing. *English Teaching Forum*, 2-5.

Hughes, A. (1989). *Testing for language teachers.* London: Cambridge University Press.

Josephson, M.I. (1989, July). "Marking" EFL compositions: A new method. *English Teaching Forum*, 28-32.

Kaplan, R. (1966). Cultural patterns in inter-cultural education. *Language Learning* 16, 1~20.

Kono, M. (1978). Language testing. In Koike, I. et al (Eds.). *The teaching of English of English in Japan* (pp. 450-463). Tokyo: Eichosha Publishing Co.

Madsen, H.S. (1983). *Techniques in testing.* New York: Oxford University Press.

Mckendy, T. (1992, May). Locally developed writing tests and the validity of holistic scoring. *Research in the Teaching of English*, 26 (2), 149-166.

Oller, J.W.Jr. (1987, October). Practical ideas for language teachers from a quarter century of language testing. *English Teaching From*, 42-46.

Raimes, A. (1985, June). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly, 19* (2), 229-258.

Saleemi, A.P. (1988, January). Language testing: Some fundamental aspects. *English Teaching Form*, 2-6.

Zamel, V. (1983, June). The composing process of advanced ESL students: Six case studies. *TESOL Quarterly, 17* (2), 165-187.